



# On Political Methodology

## Citation

King, Gary. 1991. On political methodology. *Political Analysis* 2(1): 1-29.

## Published Version

doi:10.1093/pan/2.1.1

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4313304>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

# Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# On Political Methodology

*Gary King*

## 1. Introduction

"Politimetrics" (Gurr 1972), "polimetrics," (Alker 1975), "politometrics" (Hilton 1976), "political arithmetic" (Petty [1672] 1971), "quantitative Political Science (QPS)," "governmentrics," "posopolitics" (Papayanopoulos 1973), "political science statistics" (Rai and Blydenburgh 1973), "political statistics" (Rice 1926). These are some of the names that scholars have used to describe the field we now call "political methodology."<sup>1</sup> The history of political methodology has been quite fragmented until recently, as reflected by this patchwork of names. The field has begun to coalesce during the past decade; we are developing persistent organizations, a growing body of scholarly literature, and an emerging consensus about important problems that need to be solved.

I make one main point in this article: If political methodology is to play an important role in the future of political science, scholars will need to find ways of representing more interesting political contexts in quantitative analyses. This does *not* mean that scholars should just build more and more complicated statistical models. Instead, we need to represent more of the essence of political phenomena in our models. The advantage of formal and quantitative approaches is that they are abstract representations of the political world and are, thus, much clearer. We need methods that enable us to abstract the right parts of the phenomenon we are studying and exclude everything superfluous.

---

This paper was presented at the annual meeting of the American Political Science Association, Atlanta, Georgia. My thanks go to the section head, John Freeman, who convinced me to write this paper, and the discussants on the panel, George Downs, John Jackson, Phil Schrodtt, and Jim Stimson, for many helpful comments. Thanks also to the National Science Foundation for grant SES-89-09201, to Nancy Burns for research assistance, and to Neal Beck, Nancy Burns, and Andrew Gelman for helpful discussions.

1. Johnson and Schrodtt's 1989 paper gives an excellent sense of the breadth of formal and quantitative political methods, a broad focus but still much narrower than the diverse collection of methods routinely used in the discipline. For this paper, I narrow my definition of political methodology even further to include only statistical methods.

Despite the fragmented history of quantitative political analysis, a version of this goal has been voiced frequently by both quantitative researchers and their critics (see sec. 2). However, while recognizing this shortcoming, earlier scholars were not in the position to rectify it, lacking the mathematical and statistical tools and, early on, the data. Since political methodologists have made great progress in these and other areas in recent years, I argue that we are now capable of realizing this goal. In section 3, I suggest specific approaches to this problem. Finally, in section 4, I provide two modern examples to illustrate these points.

## 2. A Brief History of Political Methodology

In this section, I describe five distinct stages in the history of political methodology.<sup>2</sup> Each stage has contributed, and continues to contribute, to the evolution of the subfield but has ultimately failed to bring sufficient political detail into quantitative analyses. For the purpose of delineating these five stages, I have collected data on every article published in the *American Political Science Review* (APSR) from 1906 to 1988.<sup>3</sup> The APSR was neither the first political science journal nor the first to publish an article using quantitative methods, and it does not always contain the highest quality articles.<sup>4</sup> Nevertheless, APSR has consistently reflected the broadest cross-section of the discipline and has usually been among the most visible political science journals. Of the 2,529 articles published through 1988, 619, or 24.5 percent, used quantitative data and methods in some way.

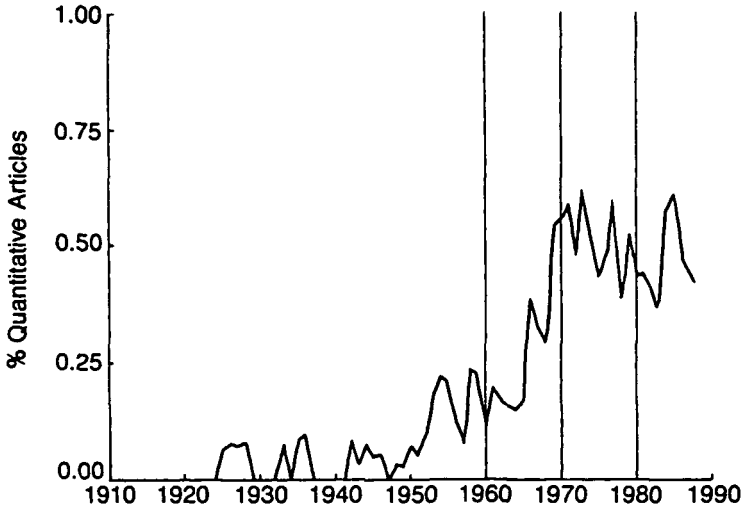
At least four phases can be directly discerned from these data. I begin by briefly describing these four stages and a fifth stage currently in progress. In these accounts, I focus on the ways in which methodologists have attempted

---

2. Although these stages in the history of political methodology seem to emerge naturally from the data I describe below, this punctuation of historical time is primarily useful for expository purposes.

3. Not much of methodological note happened prior to 1906, even though the history of quantitative analysis in the discipline dates at least to the origins of American political science a century ago: "The Establishment of the Columbia School not only marked the beginnings of political science in the U.S., but also the beginnings of statistics as an academic course, for it was at that same time and place—Columbia University in 1880—that the first course in statistics was offered in an American university. The course instructor was Richmond Mayo-Smith (1854–1901) who, despite the lack of disciplinary boundaries at the time, can quite properly be called a political scientist" (Gow 1985, 2). In fact, the history of quantitative analysis of political data dates back at least two centuries earlier, right to the beginnings of the history of statistics (see Stigler 1986; Petty, 1672).

4. Although quantitative articles on politics were published in other disciplines, the first quantitative article on politics published in a political science journal was Ogburn and Goltra 1919.



**Fig. 1. The growth of quantitative political science**

to incorporate more political context in their analyses at each juncture, and the limitations, imposed by both data and methods, to doing so effectively.

As figure 1 illustrates, political scientists first began using quantitative analysis consistently during the 1920s. This marks the first essential stage in the development of methodology. Clearly, one cannot make use of quantitative methods without data to analyze, nor model politics without systematic empirical evidence. Moreover, even before this time, scholars began to argue that hypotheses about political phenomena could and should be verified. For example, A. Lawrence Lowell wrote:

The main laboratory for the actual working of political institutions is not a library, but the outside world of political life. It is there that the phenomena must be sought. . . . Too often statements are repeated in book after book without any serious attempt at verification. (1910, 7–8)

For the most part, the first quantitative political scientists relied on direct empirical observation, as distinct from data collection. However, this proved an inadequate means of organizing and understanding the “outside world of political life.” This sphere was simply too immense to study effectively with direct observation. Thus, methodologists turned to systematic data collection, a trend that gained considerable momentum during the 1920s. Charles Merriam wrote, “Statistics increase the length and breadth of the observer’s range, giving him myriad eyes and making it possible to explore areas hitherto only

vaguely described and charted" (1921, 197). A fascinating statement about these data collection efforts can be found in a series of sometimes breathless reports by the "National Conference on the Science of Politics," instructing political scientists to collect all manner of data, including campaign literature, handbooks, party platforms in national, state, and local politics, election statistics and laws, correspondence that legislators receive from their constituencies, and many other items (see "Report" 1926, 137).

This new interest in data collection ultimately had two important consequences: First, it greatly expanded the potential range of issues that political scientists could address. Even today, political methodologists' most important contributions have been in the area of data collection; even the ICPSR was founded by political scientists.<sup>5</sup> Additionally, the availability of data naturally raised questions of how best to use it, the heart of political methodology. Although the techniques were not fully understood and widely used until many years later, political scientists first experimented with statistical techniques during this early period, including correlation, regression, and factor analysis (see Gow 1985).

Figure 1 also illustrates the second important phase in the evolution of political methodology, the "behavioral revolution" of the late 1960s. During this period, the use of quantitative methods increased dramatically. In only five years, the proportion of articles in the *APSR* using quantitative data and methods increased from under a quarter to over half. Behavioralists popularized the idea of quantification, and applied it to many new substantive areas.

Unfortunately, while the behavioralists played an important role in expanding the scope of quantitative analysis, they also contributed to the view that quantitative methods gave short shrift to political context. While innovative in finding new applications, they generally relied on methods that had been in use for decades (and still not adequately understood), applying them

---

5. Despite recent strides forward, much remains to be done. Consider one small example of shortcomings in our data, an example relatively free of theoretical complications. Evidence on incumbency advantage comes largely from two collections of aggregate election returns from U.S. Senate and House elections, one from the ICPSR and one coded by scholars from *Congressional Quarterly*. To analyze the issue of incumbency advantage properly, one needs the Democratic proportion of the two-party vote and the political party of the incumbent, if any, in each district. I compared the two collections from 1946 to 1984 and compiled a list of all House districts in which the vote proportions recorded by each differed by more than ten percent, or the incumbency code was wrong. Over these twenty elections, the number of district errors was the equivalent of nearly two full houses of Congress. Indeed, according to the ICPSR data, elections for the U.S. Senate in 1980 were held in the wrong third of the American states! Either this is the most important empirical finding in many years, or we need to pay even closer attention to issues of measurement. As Lowell once wrote, "Statistics, like veal pies, are good if you know the person that made them, and are sure of the ingredients" (1910, 10).

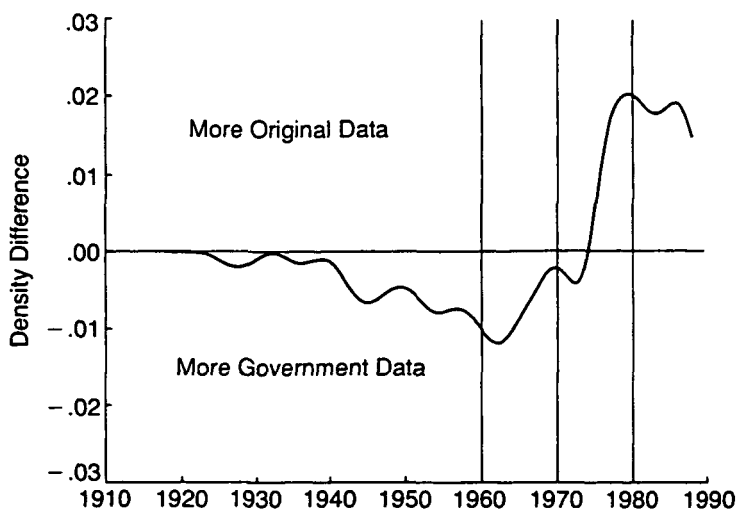


Fig. 2. Collecting data and measuring concepts

over and over again to new, if sometimes narrow, substantive questions. Thus, they encouraged the view that methods need not, or could not, be adapted to the specific problem at hand.

Figure 2 illustrates the third phase in the development of political methodology: increasing reliance on original data, rather than data automatically generated by political processes. Examples of the latter include election and roll call data. In contrast, concepts such as representation, power, and ideology require more active and creative measurement processes. Figure 2 plots the *difference* (smoothed via kernel density estimates) between articles that used existing data published by government and business and those that used more original data created by the author or other political scientists. Before the 1960s, quantitative articles relied mostly on published data from government or business (see Gosnell 1933). During the 1960s, one observes a small change in the direction of greater reliance on original data. However, this transition began in earnest during the mid-1970s, after which point almost twice as many quantitative articles used original, rather than government, data.

This sharp transition heralded an extremely important development in the history of quantitative political science. Quantitative analysts were no longer limited to areas of study for which data sets were routinely compiled. Equally important, theoretical concepts were used in designing data collection efforts. Thus, these data embodied political content that was previously quite difficult to study by systematic means. The potential scope of analysis was again

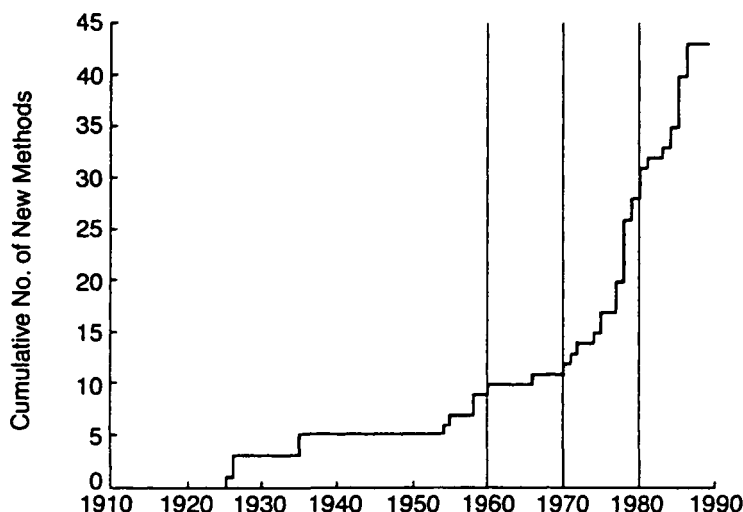


Fig. 3. New statistical methods in *APSR* articles

greatly expanded. For example, researchers used content analyses and event counts to compile data in the field of international relations, enabling them to consider such questions as the causes of war, the effects of military expenditures on the deterrence or provocation of international conflict, and of the role of specific reciprocity on the behavior of nations. Measures of variables such as international conflict, power, deterrence, and reciprocity do not naturally exist in the world, and, without the capability to create data, such questions could not be addressed systematically.<sup>6</sup>

The fourth development in the history of political methodology took place in the late 1970s, with a dramatic increase in the importation of new quantitative methods from other disciplines. Figure 3 graphs the cumulative number of new statistical methods used in *APSR*.<sup>7</sup> However, these methods were "new" only to political science. Virtually without exception, they were developed by scholars in other disciplines.

6. In the 1960s and 1970s, quantitative scholars in international relations were fond of saying that their field was like the discipline of economics in the 1950s, pausing to solve several critical measurement issues before proceeding on to bigger theoretical questions. However, as many came to realize, this was backward reasoning: The process of measuring difficult and sophisticated concepts, like those in international relations, is not fundamentally easier, or even different from, the theoretical process of deriving these concepts in the first place (see Eisner 1989).

7. Most of these methods were also new to political science, at least within one to two years of publication in *APSR*.

Borrowing from other disciplines is certainly not unique to methodology. Indeed, some of the most influential theories of political science were adapted from social psychology, economics, historical sociology, and elsewhere. The practice of importing methods has both merits and drawbacks. Importing methods provided a means of partially redressing the imbalance between our data, rich in political context, and our methods, which were not sophisticated enough to make full use of such information. For example, regression and factor analysis were introduced into political science in 1935 (see Gosnell and Gill 1935). More recently, some methods successfully imported to political science include those that allow for endogeneity (Jackson 1975), autocorrelation (Hibbs 1974) and selection bias (Achen 1986b), in regression models; such procedures have proved extremely useful in models of voting behavior and political economy, respectively. Weisberg (1974) "reimported" several methods of scaling analysis in an intuitive and influential article, and Kritzer (1978a and 1978b) introduced an easy-to-use method for coping with categorical dependent variables within the familiar regression framework. Scholars have also recently imported and developed sophisticated models of time-series (Beck 1983 and 1987; Freeman 1983 and 1989) and pooled time-series cross-sectional data (Stimson 1985), among many others.

However, precisely because they are adopted from other disciplines without substantial modification or adaptation, imported methods are sometimes ill-suited to extract all useful information from political data. An interesting example of the common problems with importing statistical methods—one of the first attempts—can be found in Rice (1926). Because his methods were so simple, the problems with imported method in this instance are transparent. Rice analyzed votes for LaFollette in midwestern states with a view to determining how attitudes and opinions diffuse. Among other things, he wished to know whether "political boundaries interpose little or no obstacle" to this diffusion. He studied this question with the method of data summaries. He used averages and standard deviations, and was especially concerned with fitting a Normal distribution to his electoral data. The latter was a practice then common among many scientists who appeared to have found evidence that "many human characteristics are distributed normally" (Rice 1926, 316–17).<sup>8</sup> Fitting the Normal distribution across counties in eight midwestern states does show how the mean and variance are interesting summary statistics, in that they conveniently describe LaFollette sentiment.<sup>9</sup> Unfortunately,

---

8. It turned out later that the statistical tests used to decide whether observed variables are distributed Normally were not very powerful. When more powerful tests were developed, scholars found extremely few examples of naturally occurring, Normally distributed variables in political science or anywhere else.

9. In modern terminology, the mean and variance are sufficient statistics for a Normal distribution.



the Normal fit itself bears virtually no relationship to the interesting political questions Rice posed, and made extremely poor use of the detailed, county-level data he obtained.

In fact, the simple method of summarizing county-level data with state-wide averages does provide some information with which to answer his interesting question, even if he did not notice this at the time. For example, Rice reports that Wisconsin counties gave LaFollette an average 54.3 percent of their votes (in part because LaFollette was a Favorite Son), whereas counties in the neighboring state of Minnesota gave him only 45.0 percent average support. A small test of Rice's question, with the data he reported, can be conducted by asking whether Minnesota's support is as high as 45.0 percent because it is next to Wisconsin, or because it contained voters with similar characteristics. The answer is uncertain, but his data do indicate that the cross-border diffusion effects he hypothesizes are not enormous: Iowa, also Wisconsin's neighbor, supported LaFollette only 28.5 percent on average, a figure lower than all but one of the other states in his sample. Had Rice focused more on studying the substance of his substantive questions, as he is otherwise famous for, he would never have wasted space on the Normal fit part of his method; at the same time, he could have focused on the other, more interesting, summary statistics enabling him to partially answer the questions he posed. This example illustrates some of the problems with imported techniques. Methodologies are not always universally applicable; they must be adapted to specific contexts and issues if data are to be put to good use.

In combination, the dramatic developments in original data collection and innovative methodology after the 1960s paint a particularly bleak picture of the behavioral revolution. During the 1960s, very little original data appeared in political science, and the learning curve of new methods in figure 3 was almost completely flat. Indeed, this was probably one of the reasons why behavioralism was so unpopular among nonquantitative researchers: Most scholars were merely applying the same methods over and over again to new areas with relatively unoriginal data, rather than developing new statistical procedures for each of the areas into which quantitative analysis was extended. Their relatively simple statistical methods guaranteed that the behavioralists could only rarely claim to have learned something about politics that could not have been learned without quantitative data. In addition, since the methods were not specially adapted for each research problem, quantification was often less than a major improvement over more traditional approaches and often looked as silly as the critics claimed. Of course, this should take away little from the critical role behavioralism played in introducing quantitative analysis to many new parts of the discipline.

In addition to these four developments in political methodology—data

collection (1920s), the growth of quantitative methods (late 1960s), measuring concepts (mid-1970s), and importing statistical techniques (late 1970s)—a fifth important development, not well represented in *APSR* data, is currently underway. In the 1980s, political methodologists have begun to solve methodological problems explicitly, evaluating and improving existing ad hoc methods of measurement, and—still too rarely—inventing new statistical methods and estimators.

This development is largely the consequence of two publications: Hanush and Jackson's (1977) textbook, which helped to enhance the level of mathematical and statistical preparedness in the discipline, and Achen's 1983 paper, which convinced many that imported statistical techniques should no longer dominate political methodology. Through the explication of two important political science problems—ecological inference and measurement error—Achen argued that we need to solve methodological problems indigenously. He encouraged methodologists to prove consistency theorems, derive confidence intervals, invent new estimators, and to be as generally creative as are other areas in political science and the methodological subfields of other disciplines.

Much recent work in methodology is directed toward these issues, beginning with Achen's own work demonstrating bias in "normal vote" estimates (1977). Beck (1986) showed how one can learn about political substance by modeling time-series processes, rather than considering these processes to be merely an estimation nuisance. Franklin (1989) proved consistency results for his "2SAIV" procedure, enabling one to estimate a regression with variables from different, independent sample surveys. Bartels (1989) derived the bias in misspecified instrumental variables models. Rivers (1988) demonstrated the bias in models of voter behavior that ignore heterogeneity of voter preferences, and Jackson (1989) derived estimators for survey data that deal effectively with small area characteristics. King demonstrated the bias and inconsistency in the way event counts have been analyzed (1988), the inefficiencies that results when representation is measured with "uniform partisan swing" (1989a), and the bias in previous measures of incumbency advantage (King and Gelman 1991), in addition to creating improved estimators in each case.

### 3. The Future of Political Methodology

I have recounted the enormous progress political methodologists have made in collecting original data and developing methods. What remains to be done? Clearly, we have not exhausted all the possibilities in these areas; vast areas of empirical work remain unexplored by sophisticated methodological analysis, particularly in the areas of comparative politics and international relations.

Numerous potentially useful techniques developed elsewhere have gone unnoticed in political science, and many interesting data sets remain to be collected. Work in these areas should obviously continue.

But what is the next logical step in the development of this subfield? I believe the answer to this question lies in our critics' complaint that political substance and quantitative analysis are incompatible. Obviously, few methodologists accept this proposition, but there is a kernel of truth in it, judging from the history of methodology: We have not done enough to integrate context in quantitative analyses. The purpose of each of the five stages delineated in the previous section was to incorporate more political substance into quantitative analyses, but these developments, even taken together, are still insufficient. Nevertheless, I believe the goal can be accomplished by making three related changes:

First, we need to relate methods more consistently and explicitly to theories of statistical inference. I have advocated the likelihood theory of inference (King 1989b), but some other approaches work well in special cases. A focus on inference will help us to distinguish models from data, and theoretical ideas from data, more clearly. Most important, statistical analyses based on well-developed theories of statistical inference can bridge the gap between theories of politics and quantitative analyses, enabling us to test theories empirically and to build upward from the data to theory. This would make political theories more generally relevant to the empirical world, and quantitative methods a more useful and integral part of political science.

Of course, statistical inference *per se* is not a panacea, and approaches based on such theories must be carefully applied. We have not always been as conscientious as we should in this regard. For example, consistency is the notion that, as the sample size tends to infinity, the sampling distribution of the estimator will collapse to a spike over the population parameter. It is a very popular and useful statistical criterion, which makes a great deal of sense when applied to data collected from sample surveys with replacement, for example. However, political methodologists have applied the consistency criterion to time-series data as well. Logically speaking, this means that we would have to wait until the end of time for the estimator to converge to the parameter, a nonsensical argument.

For another example, consider what consistency means if the analysis is based on geography, like a county in the United States. The number of counties going to infinity could mean that the landmass and number of people in the United States is increasing, but this could happen only with either a new wave of colonial domination or an expanding earth. Alternatively, we could assume that the landmass and number of people stays constant, but more counties means that fewer and fewer people are in each county. The consequence of consistency here is that, at the limit (one person per county), all

aggregation bias is eliminated. Other statistical criteria have their problems as well, and many conflict in practice, so I am not arguing that we drop consistency in particular. Clearly, we need to have open discussions of these issues and come to a consensus about which statistical criteria make sense in relation to specific substantive issues.

Along the same lines, I propose a new statistical criterion that we should consider as important as any of the more usual ones. We should ask of every new estimator: "What did it do to the data?" Statistical criteria such as consistency, unbiasedness, minimum mean square error, admissibility, etc., are all very important, particularly since, like economists (but unlike many statisticians), we tend to conceptualize our models as existing independently of the data we happen to use for estimation. However, in the end, statistical analyses never involve more than taking a lot of numbers and summarizing them with a few numbers. Knowing that one's procedures meet some desirable statistical criterion is comforting but insufficient. We must also fully understand (and communicate) just what was done to the data to produce the statistics we report. In part, this is just another call for full reporting of statistical procedures, but it is also a suggestion that we hold off using even those statistical procedures that meet the usual statistical criteria until we can show precisely and intuitively how the data are summarized. Developing estimators that are robust, adaptive, nonparametric, semiparametric, distribution free, heteroskedasticity-consistent, or otherwise unrestrictive is important, but until we clarify just what estimators like these do to our data, they are not worth using. Deriving estimators from a well-known theory of inference will help in this regard, as will the diffusion of more sophisticated mathematical and statistical background.

Second, we require more powerful statistics and mathematics to build full probabilistic models of the processes giving rise to political phenomena. Probabilistic models enable one to represent more relevant political substance in statistical models and are required by most theories of inference.<sup>10</sup>

Of course, critics claim that technical work forces out political substance, but this is precisely incorrect. Instead, the reason much quantitative political science appears so apolitical is that the relatively simple statistical procedures that are often used are incapable of representing much of the interesting political detail. The result is that many quantitative publications either have little political substance at all, or they have substance, but only in qualitative analyses that are not part of their statistical models. In either case,

---

10. For likelihood, the full probabilistic structure is used to derive parameter estimates. For methods of moments and others, only the mean and covariance matrix are used, but these moments are usually calculated from the complete probabilistic model. Bayesian inference requires all the probabilistic assumptions of a likelihood model and others to represent prior knowledge or beliefs.

the statistical analyses look superfluous to the goals of the discipline. With more sophisticated statistics and mathematics, we would be able to fine-tune our models to the unusual data and theories developed in the discipline.

Others argue against more sophisticated statistical analyses because data in political science do not meet the usual sets of statistical assumptions, but this is also completely backward. Simple statistical techniques are useful only with data of the highest quality and unambiguous content. Only with sophisticated methods will we be able to generate adequate probabilistic models of the complicated political processes giving rise to our quantitative data.

Finally, when we encounter problems for which the requisite statistical methods to incorporate all relevant political information do not yet exist, we should portray this information with descriptive statistics and graphics. Descriptive statistics as simple as sample means, standard deviations, and cross-tabulations are absent from many published works; yet, they can greatly clarify and extract useful information in quantitative data.<sup>11</sup>

Although they will clearly conflict at times, (a) closer attention to theories of inference; (b) more sophisticated stochastic modeling; and, (c) more descriptive statistics and graphics are different ways of incorporating more political substance into quantitative analyses. The gap between quantitative and qualitative scholars that is present in many departments is unlikely to be bridged without these and other steps designed for the same purpose. I turn now to an example of these points in the analysis of aggregate data.

#### 4. Aggregate Data

Virtually every subfield of political science has aggregate data of some kind, and most have analyzed them in some way. For example, scholars in American and comparative politics study electoral data and the consequences of electoral laws; in comparative politics and international relations, scholars compare nations and interactions among nations with data measured at the level of the nation-state; in international political economy, researchers seek to explain variations in indicators of the health of national economies.

The two subsections that follow illustrate the points emphasized above. In section 4.1, I give an example of a complete stochastic model of an important class of substantive problems—the process by which data are aggregated. This model may provide much of the solution to the ecological inference problem. Section 4.2 introduces the problem of spatial variation and correlation. This is an example of a set of political processes that have been

---

11. Indeed, one of the leaders in statistical graphics is a political methodologist, and his work should be incorporated more into statistical practice; see Tufte 1983; also see Cleveland 1985.

insufficiently studied in political science, but for which existing statistical models are wholly inadequate. Ultimately, probabilistic modeling is necessary to fully analyze such processes effectively. In the interim, graphical approaches can be extremely useful, as described below.

#### 4.1. Ecological Inference

In this section, I describe one way to model the aggregation process with a complete stochastic model. I believe this model (in terms of internal consistency) goes very far toward solving the chief problems with ecological inference: avoiding aggregation bias and producing realistic standard errors. Other problems such as latent variables may still remain (Achen 1983).

First, for contrast, I present the usual approach first suggested by Leo Goodman (1953). The goal is to consistently estimate the proportion of Democratic voters at time 1 who vote Democratic at time 2 ( $P$ ) and the proportion of Republican voters at time 1 who vote Democratic at time 2 ( $Q$ ), with only aggregate data. We observe  $D_1$ , the fraction of Democratic votes at time 1, and  $D_2$ , the fraction at time 2. The following is true by definition:

$$\begin{aligned} D_2 &= PD_1 + Q(1 - D_1) \\ &= Q + (P - Q)D_1 \end{aligned} \tag{1}$$

Across districts, this equation is only true if  $P$  and  $Q$  are constant. If they are not, one might think that this equation could be estimated by adding an error term,

$$D_2 = Q + (P - Q)D_1 + \epsilon, \tag{2}$$

and running a linear regression. However, an error term tacked on to the end of an accounting identity does not make a proper probabilistic model of any phenomenon. One can justify some of this procedure, in part, and fix some of the (implicit) problematic assumptions in various ways, but the result is a patchwork rather than a coherent model of the aggregation process. Unfortunately, this is also an approach that frequently yields nonsensical results in practice, such as estimated probabilities greater than one or less than zero. It also gives unrealistic assessments of the uncertainty with which the parameters are estimated.

Many scholars have written about this classic problem, and I refer readers to Achen (1981, 1983, 1986a) and Shively (1969) for the latest contributions and reviews of the literature. The key feature of the model I present below is the comprehensive probabilistic framework, enabling one to

TABLE 1. A Contingency Table

Time 1 Vote	Time 2 Vote		
	Dem	Rep	
Dem	$Y_j^D$	$n_j^D - Y_j^D$	$n_j^D$
Rep	$Y_j^R$	$n_j^R - Y_j^R$	$n_j^R$
	$Y_j$	$n_j - Y_j$	$n_j$

bring in as much of the substance of this research problem as possible. I begin with individual voters, make plausible assumptions, and aggregate up to derive a probability distribution for the observed marginals in table 1. This probability distribution will be a function of parameters of the process giving rise to the data. Brown and Payne (1986) have presented a more general form of the model discussed below, an important special case of which was introduced by McCue and Lupia (1989); Alt and King (n.d.) explicate the Brown and Payne model, derive an even more general form, provide empirical examples, and offer an easy-to-use computer program.

The object of the analysis is portrayed in table 1. The table portrays a very simple contingency table for voting at two times. As with aggregate data, the total number of Democratic and Republican voters at each time are observed:  $Y_j$ ,  $(n_j - Y_j)$ ,  $n_j^D$ , and  $n_j^R$ . The object of the analysis is to find something out about the cells of the table that remain unobserved. At time 2,  $n_j^D$  and  $n_j^R$  are observed and assumed fixed. Thus, the only randomness we must model is that leading to the realized values of  $Y_j$ , conditional on the time 1 marginals. The loyalists  $Y_j^D$  (the number of Democrats at time 1 voting Democratic at time 2) and the defectors  $Y_j^R$  (the number of Republicans at time 1 voting Democratic at time 2) are the object of this inference problem.

Begin by letting the random variable  $Y_{ij}^P$  equal 1 for a Democratic vote at time 2, and 0 otherwise, for individual  $i$  ( $i = 1, \dots, n_j^P$ ), district  $j$  ( $j = 1, \dots, J$ ), and time 1 vote for Party  $P$  ( $P = [D, R]$ ). Then define  $\Pr(Y_{ij}^P = 1) = \pi_{ij}^P$  as the probability of this individual voting Democratic at time 2 (that is, the probability of being a loyalist if  $P = D$  or defector if  $P = R$ ). By assuming that, at time 2, the Democratic versus Republican vote choice is mutually exclusive and exhaustive, we have the result that  $Y_{ij}^P$  is a Bernoulli random variable with parameter  $\pi_{ij}^P$  for each individual. This is an almost completely unrestrictive model of individuals, as virtually all of its assumptions are easily relaxed (see Alt and King n.d.).

I now aggregate individuals in two stages. First, I aggregate these unob-

served individual probabilities within districts to get the (also unobserved) cells of the contingency table; afterward I aggregate to get the marginal random variable  $Y_j$ , the realization of which is observed in each district. To begin, let  $Y_j^P = 1/n_j^P \sum_{i=1}^{n_j^P} Y_{ij}^P$ . To get a probabilistic model for  $Y_j^P$ , we must combine our model for each  $Y_{ij}^P$  and some assumptions about the aggregation process. One possibility is to assume (1) *homogeneity*, that every individual  $i$  (in district  $j$  voting for party  $P$  at time 1) has the same probability of voting for the Democrats at time 2 ( $\pi_j^P$ ), and (2) *independence*, individual vote decisions within a district are independent of one another. If these (implausible) assumptions hold, the variable  $Y_j^P$  has a binomial distribution (in the language of introductory statistics texts,  $Y_j^P$  "successes" out of  $n_j^P$  independent trials, each with probability  $\pi_j^P$ ):  $f_B(y_j^P | \pi_j^P; n_j^P)$ .

Since these aggregation assumptions are implausible, we generalize these by letting  $\pi_j^P$  be randomly distributed across individuals within district  $j$  according to a beta distribution,  $f_B(\pi_j^P | \Pi_j^P, \alpha^P)$ , with mean  $E(\pi_j^P) = \Pi_j^P$  and dispersion parameter  $\alpha^P$ . The beta distribution is a mathematical convenience, but it is also very flexible; other choices would give very similar empirical results. Furthermore, because dependence and contagion among individuals are not identified in aggregate data, adding this assumption fixes both implausible assumptions generating the binomial distribution. To combine the binomial with the beta assumption, we calculate the joint distribution of  $Y_j^P$  and  $\pi_j^P$  and then average over the randomness in  $\pi_j^P$  within district  $j$  and for a time 1 vote for Party  $P$ . The result is the beta-binomial distribution (see King 1989b, chap. 3 for details of this derivation):

$$f_{\beta B}(y_j^P | \Pi_j^P, \alpha^P; n_j^P) = \int_0^1 f_B(y_j^P | \pi_j^P; n_j^P) f_B(\pi_j^P | \Pi_j^P, \alpha^P) d\pi_j^P \quad (3)$$

In this distribution, the mean is  $E(\pi_j^P) = \Pi_j^P$ , and the dispersion around this mean is indexed by  $\alpha^P$ . If the individual cells of the contingency table were observed, this would be a very plausible model one might use to estimate the district transition probabilities  $\Pi_j^P$ , a more general one than the usual log-linear model. When  $\alpha^P \rightarrow 1$ , the assumptions of homogeneity and independence hold so that this beta-binomial distribution converges to a binomial distribution.

Since only the margins are usually observed in table 1, we need to aggregate one further step:  $Y_j = Y_j^D + Y_j^R$ . Since we now have a probability model for the each term on the right side of this expression, we need only one assumption to get the distribution for the marginal total,  $Y_j$ . The assumption I use is that, conditional on the parameters  $(\Pi_j^D, \Pi_j^R, \alpha^D, \alpha^R)$  and the time 1



marginals that are known at time 2 ( $n_j^D, n_j^R$ ),  $Y_j^D$  and  $Y_j^R$  are independent for all districts  $j$ .<sup>12</sup> The result is an aggregated beta-binomial distribution:

$$f_{ABB}(Y_j | \Pi_j^D, \Pi_j^R, \alpha^D, \alpha^R; n_j^D, n_j^R) = \int_0^1 f_{BB}(Y_j^D | \Pi_j^D, \alpha^D; n_j^D) f_{BB}(Y_j - Y_j^D | \Pi_j^R, \alpha^R; n_j^R) d\Pi_j^D \quad (4)$$

This probability distribution is a model of the randomness in the time 2 marginal total  $Y_j$ , conditional on the time 1 marginals,  $n_j^D$  and  $n_j^R$ . The distribution is a function of four very interesting parameters that can be estimated: (1) the average probability in district  $j$  of time 1 Democrats voting Democratic at time 2,  $\Pi_j^D$ ; (2) individuals' variation around this average,  $\alpha^D$ ; (3) average probability in district  $j$  of time 1 Republicans voting Democratic at time 2,  $\Pi_j^R$ ; and (4) individuals' variation around this average,  $\alpha^R$ .

If we were to regard the transition probabilities as constant, we could just drop the subscript  $j$  and use equation 4 as the likelihood function for observation  $j$ . Alternatively, we can let  $\Pi_j^D$  and  $\Pi_j^R$  vary over the districts as logistic functions of measured explanatory variables:

$$\Pi_j^D = [1 + \exp(-X, \beta)]^{-1} \quad (5)$$

$$\Pi_j^R = [1 + \exp(-Z, \gamma)]^{-1}$$

Where  $X$  and  $Z$  are vectors of (possibly different) explanatory variables, and  $\beta$  and  $\gamma$  are the effects of  $X$  and  $Z$ , respectively, on  $\Pi_j^D$  and  $\Pi_j^R$ . Thus, even though one does not observe these loyalty and defection rates directly, one could use this model to study many interesting questions. For example, with only aggregate data, we can discover whether people are more loyal to their parties in open seats than in districts with incumbent candidates.

Since we have a full stochastic model, estimation is straightforward. We merely form the likelihood function by taking the product of equation 4 over the  $j$  districts and substituting in equations 5. One can then get the maximum likelihood estimates by taking logs and maximizing the function with respect to  $\beta$ ,  $\gamma$ ,  $\alpha^D$ , and  $\alpha^R$ , given the data. Alternatively, one can calculate the mean and covariance matrix from this distribution and use the method of moments to estimate the same parameters.

The advantage of this approach is that much more of the substance of the research problem is modeled. One is not only able to infer the unobserved

12. This assumption only requires that the probabilities be sufficiently parametrized. Analogously, in regression models, the correct explanatory variables can whiten the disturbances.

transition probabilities, but, perhaps even more significantly, we can also study the variation in these probabilities across people within districts. Even the cost of not knowing the cell frequencies (the move from the beta-binomial to the aggregated beta-binomial) is made very clear by this approach because it is an explicit part of the modeling process. Perhaps the biggest advantage over previous approaches is that this full probabilistic model should produce more reasonable estimates of uncertainty. This approach requires more sophisticated mathematics than usual. We require considerable empirical testing before recommending its general application. Indeed, once the mathematics are fully understood, interpreting results from this model in terms of inferences to individual-level parameters will be considerably easier than approaches that are farther from the substantive process generating aggregated data.<sup>13</sup>

#### 4.2. Spatial Variation and Spatial Autocorrelation

The processes generating spatial variation and spatial autocorrelation begin where the model of ecological inference in section 4.1 leaves off. These models apply to two types of data. The first are ecological data, where the object of inference is still the unobserved individual data. Scholars who write about ecological inference usually argue that this is the object of analysis for virtually all aggregate data. However, a second type of data is relatively common outside of American politics—data which are most natural at the level of the aggregate. For example, national economic statistics or measures of the degree to which a nation is democratic or representative of its people do not apply well to anything but the aggregate unit. Spatial variation and autocorrelation models apply to both types of data, but I focus here only on the second type to simplify matters.

Political scientists have collected enormous quantities of aggregate data organized by location. The local or regional component is recognized, if not adequately analyzed, in most of these data. But we forget that even sample survey data have areal components. For example, the 1980 American National Election Study samples only from within 108 congressional districts. Yet, most standard models using these data ignore this feature, implicitly assuming that the same model holds within each and every congressional district.

Most other subfields of political science also pay sufficient attention to the spatial features of their data. For purposes of analysis, we often assume all districts (or countries, or regions) are independent, even though this is almost certainly incorrect. Perhaps even more problematic, we do not sufficiently

---

13. Once the mathematics are understood, the approach also meets my what-did-you-do-to-those-data criterion, since the stochastic model is quite clear.

explore spatial patterns in our data. Although maps of political relationships (usually with variables represented by shading) were once relatively common in American politics, for example, they are now almost entirely missing from this literature. Think of how much political information was represented in the classic maps in *Southern Politics*, for example, where V. O. Key (1949) graphically portrayed the relationships between racial voting and the proportion blacks. Key also showed spatial relationships by circling groups of points in scatter plots to refer to specific geographic areas.<sup>14</sup>

Statistical models have been developed to deal with both spatial variation and spatial autocorrelation, but these models are very inadequate to the task of extracting contextual political information from geographic data—far more inadequate than in other areas of statistical analysis. For example, compare the parameter estimates and standard errors from a time-series analysis to the original data: a plot over time of the complete original data may show a few years (say) that are especially large outliers. A similar comparison for spatial data is usually far different: One can plot residuals on a map and find a rich variety of geographic patterns, which may, in turn, suggest numerous other causes and relationships in the data (see Jackson 1990).

All this suggests two critical tasks for political methodologists: (1) we should find ways to improve existing statistical models along these lines, and (2) in the interim and perhaps indefinitely, we need to encourage much more attention to mapping and other similar graphical images. Although one could not emphasize the latter enough, I will spend the remainder of this section focusing on the inadequacy of statistical models of geographic data.

I first describe models of spatial variation, and then turn to models of spatial autocorrelation. In both cases, I discuss only linear-Normal models. I do this for simplicity of presentation, not because these are more generally appropriate than any other functional form or distributional assumption.

### *Spatial Variation*

Spatial variation is what we usually think of when we consider the special features of geographic data. Take, for example, the linear regression model where the unit of analysis is a geographic unit:

$$E(Y) \equiv \mu = X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots \beta_k X_k \quad (6)$$

14. If one were to argue that a classic book like V. O. Key's might just be the exception, consider Dahl's *Who Governs?* (1961). Another classic, but without a single map. Dahl could have even more vividly portrayed the nature of politics in this city by showing exactly where each of the wards he described was located, where the city hall was, and where each of the key actors lived. He does have a few graphs with wards distinguished, but without a map this political context in his quantitative data is lost. (Obviously, the book hardly needs more in the way of political context, but I am focusing only on the degree to which he showed the political context in his quantitative data.)

If  $\mu$  is the expected degree of political freedom in a country, then  $X$  can include regional variables or attributes of the countries in the sample. A plot of the fitted values and residuals on a map of the world would give one a good sense of how well this model was representing the political question at hand. In a model like this, estimates of the effect parameters,  $\beta$ , are unlikely to provide as much substantive information.

A generalization of this model that takes into account some more of the geographical information was proposed by Brown and Jones (1985). Their idea (sometimes called "the expansion method"; see Anselin 1988) was to take one or more of the effect parameters (say  $\beta_1$ ) and to suppose that it is not constant over the entire map. They let it vary smoothly as a quadratic interaction of north-south and east-west directions:

$$\beta_1 = \gamma_0 + \gamma_1 x + \gamma_2 y + \gamma_3 xy + \gamma_4 x^2 + \gamma_5 y^2 \quad (7)$$

where  $x$  and  $y$  are Cartesian coordinates (not independent and dependent variables). One can then substitute this equation for  $\beta_1$  in equation 6 to estimate  $\beta$ , and  $\gamma_0, \dots, \gamma_5$ . Finally, we can portray the results by plotting the estimated values of  $\hat{\beta}_1$  as a continuous function over a map; this can take the form of a contour plot, a three-dimensional density plot, or just appropriate shading.

This model will obviously give one a good sense of where the effect of  $X_1$  is largest, but for any interesting political analysis, equation 7 is a vast oversimplification. Why should the effect of  $X_1$  vary exactly (or even approximately) as a two-dimensional quadratic? The model also ignores political boundaries and cannot cope with other discontinuous geographic changes in the effect parameter.

A switching regression approach can model discontinuous change, but only if the number of such changes are known a priori, an unlikely situation (Brueckner 1986). One could also apply random coefficient models or other approaches, but these are also unlikely to represent a very large proportion of the spatial information in the typical set of quantitative political data.

### *Spatial Autocorrelation*

Spatial autocorrelation—where neighboring geographic areas influence each other—is an even more difficult problem than modeling spatial variation. To get a sense of the problem, begin with the set of models for time-series processes. Some of the enormous variety of time-series models can be found in Harvey 1981. In political science, Beck (1987) showed that one of the highest quality time-series in the discipline—presidential approval—provided insufficient evidence with which to distinguish among most substantively interesting time-series models. Freeman (1989) complicated matters even further when he demonstrated that the aggregation of one time-series

process can be an entirely different process. Now imagine how much more complicated these standard models would be if time travel were possible and common. This is basically the problem of spatial autocorrelation.

Geographers have tried to narrow this range of possible models somewhat with what Tobler (1979) called the first law of geography: "everything is related to everything else, but near things are more related than distant things." Unfortunately, in political science, even this "law" does not always hold. For example, although regional effects in international conflict are important, the Soviet Union probably has more of an effect on U.S. foreign policy than Canada does. Similarly, New York probably takes the lead on state policy from California more frequently than from Kansas.

Virtually all models of spatial autocorrelation make use of the concept of a spatial lag operator, denoted  $W$ .  $W$  is an  $n \times n$  matrix of weights fixed a priori. The  $i, j$  element of  $W$  is set proportional to the influence of observation  $i$  on observation  $j$ , with diagonal elements set to zero, and rows and columns summing to one; thus, the matrix need not be symmetric if influence structures are asymmetric.

A simple version of the  $W$  matrix is coded zero for all noncontiguous, and  $1/c_i$  for contiguous, regions (where  $c_i$  is the number of regions contiguous to region  $i$ ). Then multiplying  $W$  into a column vector produces the average value of that vector for the contiguous regions. For example, if  $y$  is a  $(50 \times 1)$  vector containing U.S. state-level per capita income figures, then  $Wy$  is also a  $(50 \times 1)$  vector, the first element of which is the average per capita income for all states contiguous to state 1.

Numerous models have been proposed for spatial processes, but virtually all are functions of what I call the *spatial fundamentals*: (1) explanatory variables, (2) spatially lagged dependent variables, (3) spatially lagged shocks, and additional spatial lags of each of these.<sup>15</sup> For linear models, explanatory variables are portrayed in equation 6. We can add a spatial lag of the dependent variable as follows:

$$E(Y_i | y_j, \forall i \neq j) = X_i \beta + \rho W_i y \quad (8)$$

Note that this is expressed as a *conditional* expectation, where the dependent variable for region  $i$  is conditional on its values in all other regions. The second term on the right side of this equation includes the spatial lag of the dependent variable for region  $i$  ( $W_i$  is the first row of the  $W$  matrix). The idea behind this model is that the lagged dependent variable in some geographic areas (say racially polarized voting) may influence the expected value in others.

15. This presentation is the spatial analogy to the categorization of time-series models in King 1989b, chap. 7.

Another way to think about this model is to consider the *unconditional* expectation. In time-series models, the conditional and unconditional representations are mathematically equivalent because the random variables for all times before the present are already realized and thus known. In spatial models, variable  $y$  on the right side of the equation is known only because of the conditional expectation.

To write the unconditional expectation of equation 8, we merely take the expected value of both sides of this equation and recursively reparameterize:

$$E(E(Y_i|y_j, \forall j \neq i)) = X_i\beta + \rho W_i E(y) \quad (9)$$

$$\begin{aligned} E(Y_i) &= X_i\beta + \rho W_i [X_i\beta + \rho W_i E(y)] \\ &= X_i\beta + \beta\rho(W_i X_i) + \rho^2 W_i^2 E(y) \\ &= X_i\beta + \beta\rho(W_i X_i) + \rho^2 W_i^2 [X_i\beta + \beta\rho(W_i X_i) + \rho^2 W_i^2 E(y)] \\ &= X_i\beta + \beta\rho(W_i X_i) + \beta\rho^2(W_i^2 X_i) + \beta\rho^3(W_i^3 X_i) + \dots, \end{aligned}$$

where we use the notation  $W_i^2 y$  for row  $i$  of the matrix  $WWy$ . This unconditional form also provides a very interesting substantive interpretation for the model since the expected value of  $Y_i$  is written as a geometric distributed spatial lag of explanatory variables. Thus, the first term is the effect of  $X_i$  on  $E(Y_i)$  (e.g., the effect of the proportion of blacks on racial polarization). The second term is the effect of the average values of the explanatory variables in regions contiguous to  $i$  on the dependent variable in  $i$ , after controlling for the explanatory variables in region  $i$ . (For example, racially polarized voting might be affected by the separate influences of the proportion of blacks in a county *and* in the neighboring counties; if  $0 < \rho < 1$ , the effect of blacks in contiguous counties would be smaller than the effect in the same county.) The third term in the equation represents the effects of the explanatory variables in regions two steps away—in regions contiguous to the regions that are contiguous to the current region. Each additional term represents the effects of the explanatory variables in regions farther and farther away. In this model, an explanatory variable measured in region  $i$  has a direct affect on the dependent variable in region  $i$  and, through region  $i$ , has an affect on the next region, and so on.<sup>16</sup>

16. This model can be easily estimated with maximum likelihood methods. If  $Y$  is distributed Normally, the likelihood function is proportional to a Normal distribution with mean  $(I - \rho W)^{-1} X\beta$  (another parameterization of the unconditional expected value) and variance  $\sigma^2(I - \rho W)^{-1}(I - \rho W)^{-1}$ .

Alternatively, we can write a different conditional model with explanatory variables and spatially lagged random shocks as follows:

$$\begin{aligned} E(Y_i | y_j, \forall j \neq i) &= X_i \beta + \rho W_i \epsilon_i \\ &= X_i \beta + \rho W_i (y_i - X_i \beta). \end{aligned} \quad (10)$$

Note how this model compares to the conditional model in equation 8. Both are linear functions of a vector of explanatory variables. In addition, instead of neighboring values of the dependent variable affecting the current dependent variable, this model assumes that random shocks (unexpected values of the dependent variable) in neighboring regions affect a region. For example, a reasonable hypothesis is that, after taking into account the explanatory variables, only *unexpected* levels of international conflict in neighboring countries will produce conflict in one's own country (see Doreian 1980).

The unconditional version of this model takes a surprisingly simple form:

$$\begin{aligned} E(E(Y_i | y_j, \forall j \neq i)) &= X_i \beta + \rho W_i E(\epsilon_i) \\ E(Y_i) &= X_i \beta + \rho W_i [E(y) - E(X_i \beta)] \\ &= X_i \beta. \end{aligned} \quad (11)$$

What this means is that values of the explanatory variables have effects only in the region for which they are measured. Unlike the first model in equations 8 (the conditional version) and 9 (the unconditional version), the explanatory variables do not have effects that lop over into contiguous regions in this second model in equations 10 (the conditional version) and 11 (the unconditional version). Only unexpected, or random, shocks affect the neighboring region. Once these shocks affect the neighboring region, however, they disappear; no "second-order" effects occur where something happens in one county, which affects the next county, which affects the next, etc.

A more sophisticated model includes all three spatial fundamentals in the same model (see Brandsma and Ketallapper 1978; Doreian 1982; and Dow 1984; on the "biparametric" approach):

$$E(Y_i | Y_j, \forall j \neq i) = X_i \beta + \rho_1 W_{1i} y + \rho_2 W_{2i} \epsilon. \quad (12)$$

This model incorporates many interesting special cases, including the previous models, but it is still wholly inadequate to represent the enormous variety of conceivable spatial processes. For example, I have never seen a single model estimated with social science data with more than these two

spatial parameters ( $\rho_1$  and  $\rho_2$ ) or a model with more than one conditional spatial lag.

Another difficulty is the very definition of the spatial lag operator. How does one define the "distance" between irregularly shaped spatial units? If "distance" is to mean actual mileage between pairs of U.S. states, should the measure be calculated between capital cities, largest cities, closest borders, or just 0/1 variables indicating neighborhoods (Cressie and Chan 1989)? More generally, we can also use more substantively meaningful definitions of distance, such as the proportion of shared common borders, numbers of commuters traveling daily (or migrating permanently) between pairs of states, or combinations of these or other measures (see Cliff and Ord 1973 and 1981). Choosing the appropriate representation is obviously difficult, and the choice makes an important difference in practice (Stetzer 1982). These concerns are also important because unmodeled spatial variation will incorrectly appear to the analyst as spatial autocorrelation.

However, a much more serious problem is that the  $W$  matrix is not a unique representation of the spatial processes it models. This is not the usual problem of fitting a model to empirical data. It is the additional problem of fitting the model to the theoretical spatial process. For example, begin with a spatial process, and represent it with a matrix  $W$ . The problem is that one cannot reconstruct the identical map from this matrix. Since the  $W$  matrix (and  $X$ ) is the only way spatially distributed political variables are represented in these models, this nonuniqueness is a fundamental problem. In order to get more politics into this class of models, we need to develop better, more sophisticated models and probably some other way to represent spatial information.<sup>17</sup>

Through all of these models, the same problem remains: statistical models of spatial data do not represent enough of the political substance existing in the data. I do not have a solution to this problem, but one possibility may lie in a literature now forming on the statistical theory of shape (see Kendall 1989 for a review). The motivation behind this literature is often archaeological or biological; for example, scholars sometimes want to know, apart from random error, if two skulls are from the same species. In this form, the literature has little to contribute to our endeavors (although political scientists are sometimes interested in shape alone; see Niemi et al. 1989). However, these scholars are working on ways of representing shapes in statistical models, and

---

17. Many other approaches have been suggested for these models. For example, Arora and Brown (1977) suggest, but do not estimate, a variety of more traditional approaches. Burridge (1981) demonstrates how to test for a common factor in spatial models; the purpose of this is to reduce the parametrization (just as Hendry and Mizon [1978] do in time-series models). For a comprehensive review of many models, see Anselin 1988 from a linear econometric viewpoint and Besag 1974 from a statistical perspective.



geographic shapes are just two-dimensional special cases of their models. Eventually, some kind of spatially continuous model that includes the shape of geographic areas along with information about continuous population densities across these areas may help to represent more politics in these statistical models. Until then, graphical approaches may be the only reasonable option.<sup>18</sup>

## 5. Concluding Remarks

Although the quantitative analysis of political data is probably older than the discipline of political science, the systematic and self-conscious study of political methodology began much more recently. In this article, I have argued for a theme that has pervaded the history of quantitative political science and the critics of this movement: In a word, the future of political methodology is in taking our critics seriously and finding ways to bring more politics into our quantitative analyses.

My suggestions for including more political context include using more sophisticated stochastic modeling; understanding and developing our own approach to, and perspectives on, theories of inference; and developing and using graphic analysis more often. I believe these are most important, but other approaches may turn out to be critical as well. For example, the probabilistic models I favor usually begin with assumptions about individual behavior, and this is precisely the area where formal modelers have the most experience. If our stochastic models are to be related in meaningful ways to

---

18. Another way geographic information has been included in statistical models is through "hierarchical" or "multilevel" models. This is different from time-series-cross-sectional models (see Stimson 1985; Dielman 1989). Instead, the idea is to use a cross-section or time-series *within* each geographic unit to estimate a separate parameter. One then posits a second model with these parameter estimates as the dependent variable varying spatially (the standard errors on each of these coefficients are usually used as weights in the second stage). These models have been developed most in education (see Raudenbush 1988; Raudenbush and Bryk 1986; Bryk and Thum 1988). The same problems of representing political information exist as in the previous section, but these models have an additional problem that has not even been noticed, much less been solved.

The problem is selection bias (see Achen 1986b), and it is probably clearest in education, where hierarchical models are in the widest use. For example, if schools are the aggregates, the problem is that they often choose students on the basis of expected quality, which is obviously correlated with the dependent variable at the first stage. The result is that the coefficients on the within-school regressions are differentially afflicted by selection bias. Much of the aggregate level regression, then, may just explain where selection bias is worse rather than the true effects of social class on achievement.

This problem is less severe in political data based on fixed geographic units like states (King and Browning 1987; King 1991), but one should check for problems that could be caused by intentional or unintentional gerrymandering.

political science theory, formal theory will need to make more progress and the two areas of research must also be more fully integrated.

Finally, as the field of political methodology develops, we will continue to influence the numerous applied quantitative researchers in political science. Our biggest influence should probably always be in emphasizing to our colleagues (and ourselves) the limitations of all kinds of scientific analysis. Most of the rigorous statistical tools we use were developed to keep us from fooling ourselves into seeing patterns or relationships where none exist. This is one area where quantitative analysis most excels over other approaches, but, just like those other approaches, we still need to be cautious. Anyone can provide some evidence that he or she is right; a better approach is to try hard to show that you are wrong and to publish only if you fail to do so. Eventually we may have more of the latter than the former.

#### REFERENCES

- Achen, Christopher H. 1977. "Measuring Representation: Perils of the Correlation Coefficient." *American Journal of Political Science* 21 (4): 805-15.
- Achen, Christopher H. 1981. "Towards Theories of Data." In *Political Science: The State of the Discipline*. ed. Ada Finifter. Washington, D.C.: American Political Science Association.
- Achen, Christopher H. 1983. "If Party ID Influences the Vote, Goodman's Ecological Regression is Biased (But Factor Analysis is Consistent)." Photocopy.
- Achen, Christopher H. 1986a. "Necessary and Sufficient Conditions for Unbiased Aggregation of Cross-Sectional Regressions." Presented at the 1986 meeting of the Political Methodology Group, Cambridge, Mass.
- Achen, Christopher H. 1986b. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.
- Arora, Swarnjit S., and Murray Brown. 1977. "Alternative Approaches to Spatial Correlation: An Improvement Over Current Practice." *International Regional Science Review* 2 (1): 67-78.
- Alker, Hayward R., Jr. 1975. "Polimetrics: Its Descriptive Foundations." In *Handbook of Political Science*, ed. Fred Greenstein and Nelson Polsby. Reading: Addison-Wesley.
- Alt, James, and Gary King. N.d. "A Consistent Model for Ecological Inference." In progress.
- Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.
- Bartels, Larry. 1989. "Misspecification in Instrumental Variables Estimators." Presented at the annual meeting of the Political Science Methodology Group, Minneapolis.
- Beck, Nathaniel. 1983. "Time-varying Parameter Regression Models." *American Journal of Political Science* 27:557-600.

- Beck, Nathaniel. 1986. "Estimating Dynamic Models is Not Merely a Matter of Technique." *Political Methodology* 11 (1-2): 71-90.
- Beck, Nathaniel. 1987. "Alternative Dynamic Specifications of Popularity Functions." Presented at the annual meeting of the Political Methodology Group, Durham, N.C.
- Besag, Julian. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems" [with comments]. *Journal of the Royal Statistical Society, ser. B*, 36:192-236.
- Brandsma, A. S., and R. H. Ketellapper. 1978. "A Biparametric Approach to Spatial Correlation." *Environment and Planning* 11:51-58.
- Brown, Lawrence A., and John Paul Jones III. 1985. "Spatial Variation in Migration Processes and Development: A Costa Rican Example of Conventional Modeling Augmented by the Expansion Method." *Demography* 22 (3): 327-52.
- Brown, Philip J., and Clive D. Payne. 1986. "Aggregate Data, Ecological Regression, and Voting Transitions." *Journal of the American Statistical Association* 81 (394): 452-60.
- Brueckner, Jan K. 1986. "A Switching Regression Analysis of Urban Population Densities." *Journal of Urban Economics* 19:174-89.
- Bryk, Anthony S., and Yeow Meng Thum. 1988. "The Effects of High School Organization on Dropping Out: An Exploratory Investigation." University of Chicago. Mimeo.
- Burridge, P. 1981. "Testing for a Common Factor in a Spatial Correlation Model." *Environment and Planning, A*. 13:795-800.
- Cleveland, William S. 1985. *The Elements of Graphing Data*. Monterey, Calif.: Wadsworth.
- Cliff, Andrew D., and J. Keith Ord. 1973. *Spatial Correlation*. London: Pion.
- Cliff, Andrew D., and J. Keith Ord. 1981. *Spatial Processes: Models and Applications*. London: Pion.
- Cressie, Noel, and Ngai H. Chan. 1989. "Spatial Modeling of Regional Variables." *Journal of the American Statistical Association* 84 (406): 393-401.
- Dahl, Robert. 1961. *Who Governs? Democracy and Power in an American City*. New Haven: Yale University Press.
- Dielman, Terry E. 1989. *Pooled Cross-Sectional and Time Series Data Analysis*. New York: M. Dekker.
- Doreian, Patrick. 1980. "Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects?" *Sociological Methods and Research* 9 (1): 29-60.
- Doreian, Patrick. 1982. "Maximum Likelihood Methods for Linear Models." *Sociological Methods and Research* 10 (3): 243-69.
- Dow, Malcolm M. 1984. "A Biparametric Approach to Network Correlation." *Sociological Methods and Research* 13 (2): 210-17.
- Eisner, Robert. 1989. "Divergences of Measurement and Theory and Some Implications for Economic Policy." *American Economic Review* 79 (1): 1-13.
- Franklin, Charles. 1989. "Estimation across Data Sets: Two-Stage Auxiliary Instrumental Variables Estimation (2SAIV)." *Political Analysis* 1:1-24.
- Freeman, John. 1983. "Granger Causality and Time Series Analysis of Political Relationships." *American Journal of Political Science* 27:327-58.

- Freeman, John. 1989. "Systematic Sampling, Temporal Aggregation, and the Study of Political Relationships." *Political Analysis* 1:61-98.
- Goodman, Leo. 1953. "Ecological Regression and the Behavior of Individuals." *American Journal of Sociology* 64:610-25.
- Gosnell, Harold F. 1933. "Statisticians and Political Scientists." *American Political Science Review* 27 (3): 392-403.
- Gosnell, Harold F., and Norman N. Gill. 1935. "An Analysis of the 1932 Presidential vote in Chicago." *American Political Science Review* 29:967-84.
- Gow, David John. 1985. "Quantification and Statistics in the Early Years of American Political Science, 1880-1922." *Political Methodology* 11 (1-2): 1-18.
- Gurr, Ted Robert. 1972. *Politimetrics: An Introduction to Quantitative Macropolitics*. Englewood Cliffs, N.J.: Prentice-Hall.
- Hanushek, Eric, and John Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.
- Harvey, A. C. 1981. *The Econometric Analysis of Time Series*. Oxford: Philip Allan.
- Hendry, David, and G. Mizon. 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England." *Economic Journal* 88:549-63.
- Hibbs, Douglas. 1974. "Problems of Statistical Estimation and Causal Inference in Time-Series Regression Models." *Sociological Methodology* 1974, 252-308.
- Hilton, Gordon. 1976. *Intermediate Poltimetrics*. New York: Columbia University Press.
- Jackson, John E. 1975. "Issues, Party Choices, and Presidential Voting." *American Journal of Political Science* 19:161-86.
- Jackson, John E. 1989. "An Errors-in-Variables Approach to Estimating Models with Small Area Data." *Political Analysis* 1:157-80.
- Jackson, John E. 1990. "Estimation of Variable Coefficient Models." Presented at the annual meeting of the American Political Science Association, Chicago.
- Johnson, Paul E., and Phillip A. Schrodt. 1989. "Theme Paper: Analytic Theory and Methodology." Presented at the annual meeting of the Midwest Political Science Association, Chicago.
- Kendall, David G. 1989. "A Survey of the Statistical Theory of Shape." *Statistical Science* 4 (2): 87-120.
- Key, V. O. 1949. *Southern Politics in State and Nation*. New York: Knopf.
- King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30 (3): 666-87.
- King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for The Exponential Poisson Regression Model." *American Journal of Political Science* 32 (3): 838-63.
- King, Gary. 1989a. "Representation Through Legislative Redistricting: A Stochastic Model." *American Journal of Political Science* 33 (4): 787-824.
- King, Gary. 1989b. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.
- King, Gary. 1991. "Constituency Service and Incumbency Advantage." *British Journal of Political Science*. Forthcoming.

- King, Gary, and Robert X. Browning. 1987. "Democratic Representation and Partisan Bias in Congressional Elections." *American Political Science Review* 81:1251-73.
- King, Gary, and Andrew Gelman. 1991. "Estimating Incumbency Advantage Without Bias." *American Journal of Political Science*. Forthcoming.
- Kritzer, Herbert M. 1978a. "Analyzing Contingency Tables by Weighted Least Squares: An Alternative to the Goodman Approach." *Political Methodology* 5 (4): 277-326.
- Kritzer, Herbert M. 1978b. "An Introduction to Multivariate Contingency Table Analysis." *American Journal of Political Science* 21:187-226.
- Lowell, A. Lawrence. 1910. "The Physiology of Politics." *American Political Science Review* 4 (1): 1-15.
- Merriam, Charles. 1921. "The Present State of the Study of Politics." *American Political Science Review* 15 (2): 173-85.
- Niemi, Richard G., Bernard Grofman, Carl Carlucci, and Thomas Hofeller. 1989. "Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan Gerrymandering." Photocopy.
- Ogburn, William, and Inez Goltra. 1919. "How Women Vote: A Study of an Election in Portland, Oregon." *Political Science Quarterly* 34:413-33.
- Papayanopoulos, L., ed. 1973. "Democratic Representation and Apportionment: Quantitative Methods, Measures, and Criteria." *Annals of the New York Academy of Sciences* 219:3-4.
- Petty, William. [1672] 1971. *Political Anatomy of Ireland*. Reprint. Totowa, N.J.: Rowman and Littlefield.
- Rai, Kul B., and John C. Blydenburth. 1973. *Political Science Statistics*. Boston: Holbrook Press.
- Raudenbush, Stephen W. 1988. "Educational Applications of Hierarchical Linear Models: A Review." *Journal of Educational Statistics* 13 (2): 85-116.
- Raudenbush, Stephen W., and Anthony S. Bryk. 1986. "A Hierarchical Model for Studying School Effects." *Sociology of Education* 59:1-17.
- "Reports of the National Conference on the Science of Politics." 1924. *American Political Science Review* 18 (1): 119-48.
- "Report of the Committee on Political Research." 1924. *American Political Science Review* 18 (3): 574-600.
- "Reports of the Second National Conference on the Science of Politics." 1925. *American Political Science Review* 19 (1): 104-10.
- "Report of the Third National Conference on the Science of Politics." 1926. *American Political Science Review* 20 (1): 124-39.
- Rice, Stuart A. 1926. "Some Applications of Statistical Method to Political Research." *American Political Science Review* 20 (2): 313-29.
- Rivers, Douglas. 1988. "Heterogeneity in Models of Electoral Choice." *American Journal of Political Science* 32 (3): 737-57.
- Shively, Philip. 1969. "Ecological Inference: The Use of Aggregate Data to Study Individuals." *American Political Science Review* 63:1183-96.
- Stetzer, F. 1982. "Specifying Weights in Spatial Forecasting Models: The Results of Some Experiments." *Environment and Planning* 14:571-84.

- Stigler, Stephen. 1986. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, Mass.: Harvard University Press.
- Stimson, James A. 1985. "Regression in Time and Space." *American Journal of Political Science* 29 (4): 914-47.
- Tobler, W. 1979. "Cellular Geography." In *Philosophy in Geography*, ed. S. Gale and G. Olsson. Dordrecht: Reidel.
- Tufte, Edward R. 1969. "Improving Data Analysis in Political Science." *World Politics* 21 (4): 641-54.
- Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. New Haven: Graphics Press.
- Weisberg, Herbert. 1974. "Dimensionland: An Excursion into Spaces." *American Journal of Political Science* 18:743-76.